Australian Government
Department of Defence
Defence Science and
Technology Organisation

Camera Network Topology Discovery Literature Review

Dmitri Kamenetsky

Intelligence, Surveillance and Reconnaissance Division

Defence Science and Technology Organisation

DSTO–GD–0667

ABSTRACT

We present a literature review on the problem of camera network topology discovery, focussing on its two main types: overlapping and non-overlapping camera fields of view. These two problems are fundamentally different and each requires a specifically tailored approach. We describe the most popular approaches for each problem and analyse their suitability for our project.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **NOV 2011**	2. REPORT TYPE	3. DATES COVERED **00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE **Camera Network Topology Discovery Literature Review**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Defence Science and Technology Organisation,PO Box 1500,Edinburgh, South Australia 5111, Australia,**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
We present a literature review on the problem of camera network topology discovery, focussing on its two main types: overlapping and non-overlapping camera fields of view. These two problems are fundamentally di erent and each requires a specifically tailored approach. We describe the most popular approaches for each problem and analyse their suitability for our project.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**15**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Published by

DSTO Defence Science and Technology Organisation
PO Box 1500
Edinburgh, South Australia 5111, Australia

Telephone: (08) 7389 5555
Facsimile: (08) 7389 6567

AR No. AR 015-219
November 2011

APPROVED FOR PUBLIC RELEASE

Camera Network Topology Discovery Literature Review

Executive Summary

The INVS (Intelligent Networked Video Surveillance) project has been ongoing in the ISRD IAE Group since 2007. The 2010-2011 year of the project focuses on multi-camera tracking and matching. The aim is to demonstrate a live camera handover system for a network of about 10 cameras with both overlapping and non-overlapping views. Hence an important decision in this project is how to handle camera network topology discovery.

Here we present a literature review on the problem of camera network topology discovery, focussing on its two main types: overlapping and non-overlapping camera fields of view. These two problems are fundamentally different and each requires a specifically tailored approach. We describe the most popular approaches for each problem and analyse their suitability for this project.

THIS PAGE IS INTENTIONALLY BLANK

Author

Dmitri Kamenetsky
Intelligence, Surveillance and Reconnaissance Division

Dmitri Kamenetsky obtained a Bachelor of Science with first class Honours in Computer Science from University of Tasmania in 2005. He then completed a PhD in statistical machine learning at the Australian National University and NICTA in 2009. His PhD thesis investigated methods for inference and parameter estimation in graphical models, in particular the Ising model. In 2010, he joined the Intelligence, Surveillance and Reconnaissance Division of the Defence Science and Technology Organisation. His research focuses on multi-camera tracking and cross-matching in video.

THIS PAGE IS INTENTIONALLY BLANK

Contents

THIS PAGE IS INTENTIONALLY BLANK

1 Introduction

Cameras are becoming cheaper and smaller day by day. There are cameras in most cell phones, surveillance cameras in subway stations, busy streets and shopping malls. There are even cameras on satellites and trucks.[1]

A *camera network* is a set of cameras (ten or as many as hundreds) monitoring some environment for a particular purpose. Camera networks may be the best way to obtain time-critical information in situations where the safety of human lives is at stake, such as, terrorist attacks in a busy subway/airport, natural disaster sites or urban combat zones. Camera networks will be essential for 21st century military, enviromental and surveillance applications [Devarajan, Cheng & Radke 2008].

Computer networks pose several research challenges to the direct application of traditional computer vision algorithms. Firstly, computer networks usually contain tens to hundreds of cameras, which is many more than is considered in many vision applications. These cameras are likely to be spread over a wide geographical area. Until recently, research on computer networks was conducted in controlled environments, where the cameras were fixed and their relation to each other was known. Nowadays it is assumed that cameras may be moved intentionally or accidentally (by being bumped) and that their configuration is not known.

Manual inspection is an inefficient and unreliable way to monitor large camera networks, especially when one needs to follow a moving target in a crowded scene. In response to this, several systems have been developed to automate the inspection task. A key part of any camera network is to understand the spatial relationships between cameras in the network. In early surveillance systems, this information was manually specified or derived during camera calibration. This process is error prone and time consuming. Furthermore, it is not robust as cameras may go down or get moved during the observation period. Hence automatic network topology discovery methods are required.

2 Topology Estimation

There are two main types of topology estimation problems [Radke 2010]: *overlapping* and *non-overlapping*. In the overlapping problem it is assumed that the cameras observe parts of the same environment from different perspectives. The relationship between the cameras can be modeled as an undirected graph, called the *vision graph* (see Section 2.1). The vision graph contains an edge between two cameras if they observe some (or all) of the same scene.

In the non-overlapping problem it is assumed that no two cameras observe the same part of the environment. Relationships between the cameras are induced by the likelihood that an object in one camera appears in another after some amount of time. These relationships can be modeled with an undirected graph called the *communication graph*, where edge weights correspond to transition probabilities and times (see Section 2.2).

The vision and communication graphs (and their computation) are fundamentally different. Consider the hypothetical network of ten cameras in Figure 1 [Devarajan, Cheng & Radke 2008].

[1] Google Maps trucks record 360° video as they drive around cities.

The presence of an edge in the communication graph does not imply the presence of the same edge in the vision graph, since nearby cameras may be pointed in different directions (*e.g.,* cameras A and C). Similarly, cameras looking at the same scene may be physically distant (*e.g.,* cameras C and F).

Some researchers do not construct the vision nor communication graphs explicitly. Instead, they model the transition probabilities and expected transition times between different regions of camera views. Such models allow one to derive both the vision and the communication graphs (see Section 2.3).

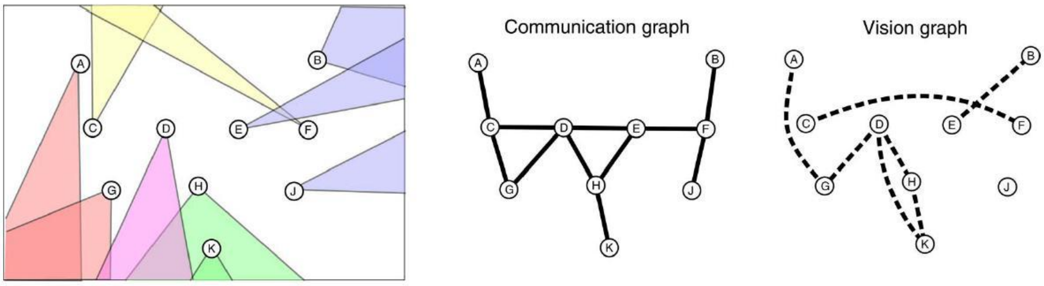

Figure 1: *(a) A camera network with ten cameras (A-K) and their corresponding fields of view. (b) The associated communication graph. (c) The associated vision graph.*

2.1 Vision Graph Estimation

Mandel, Shimshoni & Keren [2007] describe a simple algorithm for estimating the vision graph based on motion detection. The authors apply the sequential probability ratio test (SPRT) to accept/reject the possibility that two cameras observe the same scene based on the correspondence (or lack thereof) of the aggregated detections. Although the algorithm is simple, the method is not fully validated as it is only tested on a toy 3-camera network.

Van Den Hengel et al. [2007] begin by assuming that the vision graph is fully connected. Edges that contradict observed evidence are removed. In particular, an edge between two regions is removed if one camera observes movement while the other does not (in the same moment in time). The authors describe an efficient algorithm that learns the near-optimal topology of a large 100-camera network in just one hour of footage. The algorithm is robust as it is not affected by varying lighting conditions, camera angles and size of moving objects. Some examples of matching camera views are shown in Figure 2.

Devarajan, Cheng & Radke [2008] estimate the vision graph by matching features across camera views. First, each camera detects a set of distinctive feature points in its image that are likely to match other images of the same scene. Both the number of features and the length of each feature descriptor are summarised in a fixed-length structure called a *feature digest*. Each camera matches its own feature digest with those of other cameras. An edge in the vision graph is established if enough matches are found. The feature digest of an image is based on the popular and successful scale-invariant feature transform (SIFT) detector/descriptor proposed by Lowe [2004]. In particular, the feature digest is a compressed subset of SIFT features that are both distinctive and spatially

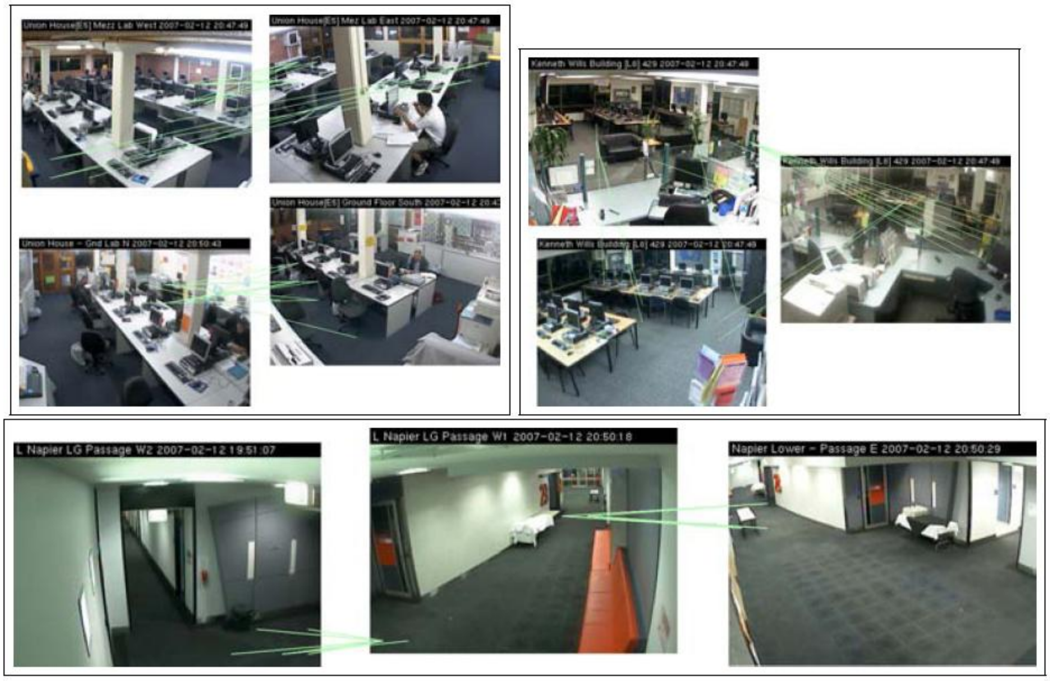

Figure 2: *Groups of images whose views are overlapping. The edges of the vision graph are shown in green.*

well distributed across the image. The paper analyzes the tradeoffs between the size of the feature digest, the number of transmitted features, the level of compression and the overall performance of edge generation.

2.2 Communication Graph Estimation

Marinakis & Dudek [2006] use a stochastic version of the Expectation-Maximization algorithm to learn plausible agent trajectories. The approach uses only detection events from the deployed sensors (equivalent to motion detection in video). The model assumes that the network is traversed by a fixed number of agents (up to 10), which travel between sensor nodes, as well as, an external sink node. Results obtained from simulations and experimental data suggest that the technique produces accurate topology graphs under a variety of conditions and compares well to other approaches.

Marinakis, Giguère & Dudek [2007] present a simple topology estimation algorithm that relies purely on the order of detection events in each camera (rather than their timestamps). The key idea is to find the smallest graph that successfully explains the observed data. Assuming that there are N agents in the environment, the algorithm considers all possible trajectories (paths) that could be taken by these agents. It then constructs the smallest graph that can explain every such path. Interestingly, when the problem is formulated in this way it is equivalent to set-covering, which is known to be NP-complete. The authors show that a simple greedy heuristic works well, even better than a more sophisticated statistical approach. The algorithm is accurate on small simulated

networks. However, no real experiments are performed and problems arise when N is smaller than the true number of agents.

2.3 Combined Graph Estimation

Makris, Ellis & Black [2004] exploit temporal correlations in observations of agents' movements through the network. The authors use Expectation-Maximization (EM) to learn a Gaussion Mixture Model (GMM) that models links between entry/exit zones. For each camera view, a set of entry/exit zones is automatically learnt [Makris & Ellis 2003].[2] A cross-correlation value is computed for each possible link from an exit zone i to an entry zone j. If the cross-correlation has a clear peak then there is a real link between i and j. Additionally, estimates of the transition times and probabilities can be extracted from the cross-correlation. If a link is detected between the zones of two cameras, then the two cameras are either adjacent (in the communication graph) or overlapping (edge in the vision graph). In particular, the two cameras are overlapping if the transition time is approximately zero; otherwise the target moves through an unseen path and so the cameras are adjacent in the communication graph. The experimental results look promising.

Correlation is effective for monotonic relationships in general, but is not flexible enough to handle multi-modal distributions. Such relationships can occur, for example, when both cars and pedestrians are part of the observations. In general, the more dense the observations and the longer the transition time, the more false correspondences will be generated by the method. With this in mind, Tieu, Dalley & Grimson [2005] improve the approach of Makris, Ellis & Black [2004]. They use more flexible, multi-modal transition distributions, and explicitly handle correspondence. This is accomplished by using mutual information as a (more general) measure of statistical dependence to estimate object correspondence. The approach makes few assumptions and does not require supervision.

3 Avoiding Topology Estimation

Camera network topology estimation is a non-trivial task that remains unsolved for a large number of cameras and complex activity patterns, such as those in a crowded public scene. Recently Wang, Tieu & Grimson [2010] proposed a method which bypasses the topology inference and correspondence problem. They use Latent Dirichlet Allocation (LDA)[3] to cluster trajectories into activities and model paths commonly taken by objects across multiple camera views. The method has few restrictions on the camera topology, the structure of the scene and the number of cameras. Evaluation is performed on two large real data sets, each of which contains more than 14,000 trajectories. On the downside, the method is limited to learning relationships among activity patterns; any temporal relationships are not discovered automatically, instead they are determined by a pre-defined temporal threshold.

Loy, Xiang & Gong [2009*a*] model the dependencies between activities across camera views with a time delayed probabilistic graphical model (TD-PGM). The nodes of the graphical model

[2] Naive K-means is used to cluster starting and ending points of object trajectories into entry/exit regions [Makris & Ellis 2003].

[3] LDA is a new standard for document analysis. The model uses multivariate beta distributions to model the relationships between words, documents and topics.

represent activites in different semantically decomposed regions from different camera views, while its directed edges encode causal relationships between these activities. The proposed approach is effective in a 9-camera network installed in a busy train station with complex and diverse scenes, such as, long queues at a ticket office, concourse, train platforms and escalators (see Figure 3). Incredibly the method works on low-quality 320×230 video running at a mere 0.7 frames per second. Note that the method is rather complex, requiring a two-stage structure learning algorithm. A simpler algorithm solving the same task is found in their earlier paper [Loy, Xiang & Gong 2009*b*].

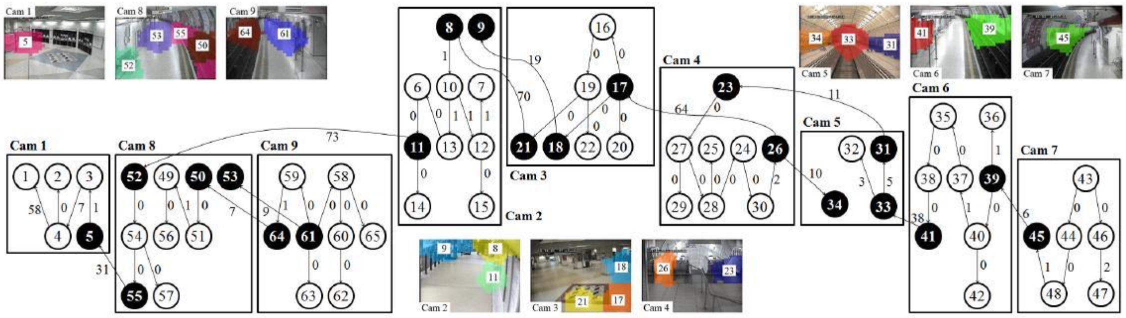

Figure 3: *The learned activity dependency graph. Edges are labeled with their associated time delays. Regions and nodes with discovered inter-camera dependencies are highlighted.*

4 Discussion

The aim of the INVS project for the year 2010-2011 is to demonstrate a live camera handover system for a network of about 10 cameras with both overlapping and non-overlapping views. To this end, we believe there are three possible options for this project:

1. Model both the vision and communication graphs independently.
2. Model a combined graph (see Section 2.3) that allows one to derive the vision and communication graphs.
3. Avoid topology estimation entirely (see Section 3).

As we move down the above list, the robustness and the generality of the methods increase. However, there is a price to pay — the methods become considerably more complex, especially in option 3.

As far as option 1 is concerned, the construction of the two graphs differ widely. The construction of the vision graph can be considered simpler as it can be solved by matching features (*e.g.,* movement, SIFT) across camera views. The communication graph, on the other hand, requires one to use sophisticated tools to model transition probabilities and time delays.

Overall it seems that option 2 is best as it achieves a good balance between method generality and implementation complexity.

References

Devarajan, D., Cheng, Z. & Radke, R. (2008) Calibrating distributed camera networks, *Proceedings of the IEEE* **96**(10), 1625–1639.

Lowe, D. G. (2004) Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* **60**, 91–110.

Loy, C. C., Xiang, T. & Gong, S. (2009*a*) Modelling activity global temporal dependencies using time delayed probabilistic graphical model, *in IEEE International Conference on Computer Vision*, pp. 120–127.

Loy, C. C., Xiang, T. & Gong, S. (2009*b*) Multi-camera activity correlation analysis, *in IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1988–1995.

Makris, D. & Ellis, T. (2003) Automatic learning of an activity-based semantic scene model, *in Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance.*, IEEE Computer Society, Los Alamitos, CA, USA, pp. 183–188.

Makris, D., Ellis, T. & Black, J. (2004) Bridging the gaps between cameras, *in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 205–210.

Mandel, Z., Shimshoni, I. & Keren, D. (2007) Multi-camera topology recovery from coherent motion, *in First ACM/IEEE International Conference on Distributed Smart Cameras*, pp. 243–250.

Marinakis, D. & Dudek, G. (2006) A practical algorithm for network topology inference, *in Proceedings of IEEE International Conference on Robotics and Automation.*, pp. 3108 –3115.

Marinakis, D., Giguère, P. & Dudek, G. (2007) Learning network topology from simple sensor data, *in Proceedings of the 20th conference of the Canadian Society for Computational Studies of Intelligence on Advances in Artificial Intelligence*, Springer-Verlag, Berlin, Heidelberg, pp. 417–428.

Radke, R. J. (2010) A survey of distributed computer vision algorithms, *in Handbook of Ambient Intelligence and Smart Environments*, Springer US, pp. 35–55.

Tieu, K., Dalley, G. & Grimson, W. (2005) Inference of non-overlapping camera network topology by measuring statistical dependence, *in Tenth IEEE International Conference on Computer Vision.*, Vol. 2, pp. 1842–1849.

Van Den Hengel, A., Dick, A., Detmold, H., Cichowski, A. & Hill, R. (2007) Finding camera overlap in large surveillance networks, *in Proceedings of the 8th Asian conference on Computer vision - Volume Part I*, Springer-Verlag, Berlin, Heidelberg, pp. 375–384.

Wang, X., Tieu, K. & Grimson, W. E. L. (2010) Correspondence-free activity analysis and scene modeling in multiple camera views, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**, 56–71.

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION DOCUMENT CONTROL DATA				1. CAVEAT/PRIVACY MARKING
2. TITLE Camera Network Topology Discovery Literature Review		3. SECURITY CLASSIFICATION Document (U) Title (U) Abstract (U)		
4. AUTHOR Dmitri Kamenetsky		5. CORPORATE AUTHOR Defence Science and Technology Organisation PO Box 1500 Edinburgh, South Australia 5111, Australia		
6a. DSTO NUMBER DSTO–GD–0667	6b. AR NUMBER AR 015-219	6c. TYPE OF REPORT General Document	7. DOCUMENT DATE November 2011	
8. FILE NUMBER 2011/1159160/1	9. TASK NUMBER NS 07/114	10. TASK SPONSOR DSTO	11. No. OF PAGES 6	12. No. OF REFS 13
13. URL OF ELECTRONIC VERSION `http://www.dsto.defence.gov.au/publications/scientific.php`		14. RELEASE AUTHORITY Chief, Intelligence, Surveillance and Reconnaissance Division		
15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT *Approved for Public Release* OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SOUTH AUSTRALIA 5111				
16. DELIBERATE ANNOUNCEMENT No Limitations				
17. CITATION IN OTHER DOCUMENTS No Limitations				
18. DSTO RESEARCH LIBRARY THESAURUS Cameras, Surveillance, Defence				
19. ABSTRACT We present a literature review on the problem of camera network topology discovery, focussing on its two main types: overlapping and non-overlapping camera fields of view. These two problems are fundamentally different and each requires a specifically tailored approach. We describe the most popular approaches for each problem and analyse their suitability for our project.				